Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 14244.2-M	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD A051537
4. TITLE (and Subtitle) A Comparative Study of Models for Reliability Growth		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Henry Braun J. Macon Paine		8. CONTRACT OR GRANT NUMBER(s) DAAG29 76 G 0298
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Princeton, New Jersey 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE July 1977
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reliability growth
Piecewise constant Poisson process
Nonhomogeneous Poisson process
Duane Model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The deterministic Duane Model as well as several stochastic models for reliability growth are studied. Graphical and quantitative techniques for assessing the performance of these models are introduced and are illustrated on a set of data. A small-scale simulation is carried out in which the effect of two factors (the number of failure modes and the range of failure intensities) are investigated. Some practical conclusions are drawn and the possibility of employing doubly stochastic Poisson Processes is discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A COMPARATIVE STUDY OF MODELS FOR RELIABILITY GROWTH

by

Henry Braun
Department of Statistics
Princeton University

J. Macon Paine*
Department of Statistics
Princeton University

Technical Report No. 126, Series 2
Department of Statistics
Princeton University
July 1977

Research sponsored by a contract with U.S. Army Research Office No. DAAG29-76-G-0298, awarded to the Department of Statistics, Princeton University.

* Currently at Bell Telephone Laboratories, Murray Hill, N.J.

ABSTRACT

The deterministic Duane Model as well as several stochastic models for reliability growth are studied. Graphical and quantitative techniques for assessing the performance of these models are introduced and are illustrated on a set of data. A small-scale simulation is carried out in which the effect of two factors (the number of failure modes and the range of failure intensities) are investigated. Some practical conclusions are drawn and the possibility of employing doubly stochastic Poisson Processes is discussed.

Key Words: Reliability growth
Piecewise constant Poisson process
Nonhomogeneous Poisson process
Duane Model

1. INTRODUCTION AND SUMMARY

In this paper we carry out a comparative study of various models for reliability growth data. We assume the reader is generally familiar with the problem. For a review of current work in the area see the AMSAA Reliability Growth Symposium [1974].

Of these models, one is deterministic and the rest stochastic. The former is due to J.T. Duane [1964] and is widely used in practice. It seemed of some interest to probe the relative strengths and weaknesses of the two approaches and, indeed, interesting differences were found.

Section 2 describes the models and discusses their motivation. In Section 3 these models are applied to a set of reliability growth data and various goodness-of-fit procedures, both graphical and quantitative, are introduced and applied. Section 4 provides the details of a Monte Carlo study that was carried out to shed more light on how well the different models performed under various conditions. Section 5 analyzes the results of the Monte Carlo, while some tentative conclusions and prospects for further study are given in Section 6.

2. THE MODELS

We first require some notation. If a device has been on test for time t and has experienced $N(t)$ failures up to that time, then $Y(t)=t/N(t)$ is called the cumulative mean time between failure (CMTBF) at time t. If the failures occurred at times

$0=t_0 < t_1 < t_2 < \ldots < t_n = T$, then we define $X_i=t_i-t_{i-1}$ $(i=1,2,\ldots,n)$ as the n interfailure times. At time point t_{i-1}, EX_i can be interpreted as the instantaneous mean time between failures (IMTBF). Contract specifications are often phrased in terms of the CMTBF reaching some assigned level. Nonetheless, the IMTBF is rather important because it gives the reliability engineer a good idea of current progress and how long until the CMTBF goal is reached. This brings us to the cruical point: The success of the models must be judged on how well they allow prediction of future failures.

The deterministic model was first introduced by J.T. Duane [1964]. He found empirically that plots of CMTBF versus operating time were approximately linear on a log-log scale. That is,

$$\ln Y(t) \mathrel{\dot{\sim}} a+b \ln t. \tag{1}$$

Reliability growth occurs when $Y(t)$ increases with t, so that the slope b is positive. To calculate the IMTBF, one computes the current "intensity" and takes reciprocals:

$$\lambda(t) = \frac{dN(t)}{dt} \sim \frac{d}{dt}\left(e^{-a}\, t^{1-b}\right) = (1-b)\, e^{-a}\, t^{-b}, \tag{2}$$

or

$$\text{IMTBF} \sim \frac{e^a \; t^b}{1-b} \; . \tag{3}$$

Of course these expressions ought not be given a stochastic interpretation. Duane suggested using least squares to estimate a and b in (1). Despite the obvious theoretical objections (after all, the successive observations on $Y(t)$ are very much dependent!) the procedure has worked well enough in practice to achieve wide success.

A stochastic model for reliability growth was proposed by L.H. Crow [1975]. Using (2) as a guide, he suggested modelling the process as a nonhomogeneous Poisson process (NHPP) with an intensity function of the form

$$\lambda(t) = \gamma\beta t^{\beta-1}, \tag{4}$$

where reliability growth is said to occur if $\beta<1$. Maximum likelihood estimates of γ and β were derived as well as formulas for CMTBF and IMTBF:

$$EY(t) = \gamma^{-1} \; t^{-\beta} \tag{5}$$

and

$$EX_i = (\gamma\beta)^{-1} \; t_{i-1}^{1-\beta} \quad (i=2,3,\ldots,n) \tag{6}$$

$$EX_1 = \gamma^{-1/\beta} \; \Gamma(1/\beta+1).$$

Note that since the intensity function varies with time, the successive interfailure times are not independent.

A second stochastic model, proposed by Braun [1976] took a different tack. One can think of reexpressing the RHS of (2) in terms of N(t) rather than t, obtaining something of the form

$$\lambda(t) = \gamma\beta\,[N(t)]^{1-1/\beta}, \tag{7}$$

and reliability growth occurs if $\beta<1$. Thus we think of the failures generated by a Poisson Process whose intensity function changes only with the number of failures, rather than with time. As a result the interfailure times X_i are independent and exponentially distributed with parameters

$$\lambda_i = \lambda(t_i) = \gamma\beta\, i^{1-1/\beta} \quad (i=1,2,\ldots,n)\ . \tag{8}$$

Of course then

$$EX_i = \frac{1}{\lambda_i} \tag{9}$$

and

$$EY(t_i) = \frac{1}{i}\sum_{j=1}^{i}\frac{1}{\lambda_j}. \tag{10}$$

A variant of this model is obtained by changing (8) so that λ_i has a different functional dependence on i:

$$\lambda_i = c \exp\left\{b\left(1-\frac{i-1}{n}\right)^2\right\} \quad i=1,2,\ldots,n, \qquad (11)$$

reliability growth occuring if $b<0$.

This has an advantage of allowing the intensities to tend to a non-zero constant, a proposal which seems quite reasonable. The parameters of these models can be easily estimated by maximum likelihood.

The estimation theory for the model(4) is given in Crow [1975], while that for models (7) and (11) can be found in Appendix A below.

The four models described above will be denoted by D,C,B1,B2 respectively. The latter three despite their somewhat ad hoc parametrizations seem reasonable competitors of model D, though they in no way attempt to model the physical processes generating the failures. We shall see in later sections how successfully they accomplish their task. Another approach to the problem would be to use techniques analogous to isotonic regression. This would involve modelling the intensity function as a step function constant between failures and required to be non-increasing. This is discussed in a report by Donelson [1975]. It may well be that this non-parametric approach could then be followed up by a parametric one, but this was not tried here. It should be noted that the question of whether to use $N(t)$ or t as the independent variable in stochastic modelling of this kind is not new. See for example the comments of T. Lewis in the discussion of a paper by D.R. Cox [1955].

3. ANALYZING THE DATA

The data set to be analyzed here consists of the recorded failure times, during development testing, of a complex electronic system. The data, which is presented in Appendix B, is composed of 52 failures which are indentified either as having been generated by one of 12 system modes (or components) or as a so-called non-pattern failure. This set is denoted by GE1. A subset of GE1, called GE2, can be derived by only recording the first failure attributable to a particular failure mode plus all the non-pattern failures. GE2 has 27 failures. Table I has the parameter estimates for all the models for both data sets. One can see that there is agreement among the models that GE2 displays the greater reliability growth.

We now proceed to consider some descriptive goodness-of-fit techniques that can be used to compare the suitability of the models, with the exception of Model D. They are all based on a probability integral transform of the interfailure times. If X_i has cdf F_i, then $F_i(X_i)$ is distributed as a uniform (0,1) deviate. In our case, F_i involves certain parameters which must be estimated from the data. As a result, the collection $\left\{\hat{F}_i(X_i)\colon i=1,2,\ldots,n\right\}$ will be only approximately uniform and independent. Nonetheless we can try plotting $\hat{F}_i(X_i)$ against i

to see whether there is any pattern or whether the $\hat{F}_i(X_i)$ are fairly well scattered about the line at .5. (This was suggested by J.W. Tukey). It is also worthwhile having a Q-Q plot of the ordered $\hat{F}_i(X_i)$ against $i/n+1$, though this should tend to look "super-regular". A quantitative feeling for the information contained in such a graph can be obtained by the following strategy.

If we let $Z_i = \hat{F}_i(X_i)$ $(i=1,2,\ldots,n)$ and compare the empirical distibution of the $\{Z_i\}$ with that of the uniform distribution, then it is well know (cf Durbin [1973] for example) that the usual goodness-of-fit statistics such as the Kolmogorov-Smirnov do not have the same distribution, even asymptotically, that they do when the true F_i is used in place of $\hat{F}_i$. Now it seems plausible that if we were to randomly partition the data into groups each containing say 10% of the points, and carry out a goodness-of-fit test separately for each group, then the effect of having estimated parameters from the whole data set should be small, when considered from the point of view of a single group.

In fact, we can use the methods of simulataneous inference to obtain an approximate overall level of significance for this procedure. A theoretical justification of this approach is currently under preparation by the first author.

Crow [1975] has adapted the work of Darling [1955] and shown how to make use of the Cramier-von Mises statistic for testing the adequacy of Model C with parameters estimated. He has shown that if the model is correct, then the distribution of the statistic is independent of the actual values assumed by the parameters and has by Monte Carlo computed the critical values of the statistic. Note that this approach works only for Model C, whereas the method suggested previously is quite general in its applicability.

Figure 1 contains a plot of $\hat{F}(X_i)$ versus i for Model B1 with GE1 data and Figure 2 contains a plot of $\hat{F}(X_i)$ versus i for Model C with GE2 data. These plots as well as others not shown here are reasonably patternless, having a hint of a positive slope which would suggest that the (stochastic) models tend to overestimate the earlier failure times and underestimate the later ones. The *Q-Q* plots are all well behaved. The modified Kolmogorov-Smirnov statistic was computed for each case and was never found significant (α=.05). In addition, the modified Cramier-von Mises statistics was calculated for Model C and was not significant (α=.10). Finally, Figure 3 plots ln Y(t) vs ln t for the GE1 data. Note that it is a bit u-shaped, indicating that CMTBF does not increase until the process is well under way.

What have we learned? Only that both data sets exhibit reliability growth, GE2 more so than GE1, and that the models seem

to do an equally good job of fitting the data. In fact, our graphical and quantitative techniques so far give us little reason to choose between the models. It would be worthwhile to see whether the forms of the different models have consequences which can be used in an exploratory (graphical) way.

Suppose we let E_i denote the expected value of X_i, the i^{th} interfailure time. Then under Model B1,

$$\frac{E_{i+1}}{E_i} = \left(\frac{i}{i+1}\right)^{1-1/\beta},$$

or

$$\log \frac{E_{i+1}}{E_i} \mathrel{\dot{\sim}} -(1-1/\beta)\left(\frac{1}{i}\right). \tag{12}$$

Under Model C, on the other hand,

$$\frac{E_{i+1}}{E_i} = \frac{\Gamma(1/\beta+i)}{\Gamma(i+1)} \cdot \frac{\Gamma(i)}{\Gamma(1/\beta+i-1)}$$

So approximately,

$$\frac{E_{i+1}}{E_i} \mathrel{\dot{\sim}} \left(\frac{1}{\beta}\right)\left(\frac{1}{i}\right) + \frac{i-1}{i}. \tag{13}$$

Of course, (13) can be obtained approximately from (12) by expansion of the logarithm. Nonetheless it may be that one approach

is better in dealing with real data. Consider now replacing expected values in (12) and (13) by the actual observations. If Model B1 holds, then plotting log (X_{i+1}/X_i) against $1/i$ should yield a linear regression through the origin with slope $-(1-1/\beta)$. In fact the estimated slope could be used to obtain an estimate of β. If Model C holds, then plotting X_{i+1}/X_i against $1/i$ should yield a linear regression. This plan was carried out with interesting, but by no means conclusive results.

Using the GE1 data as a base, the interfailure times $\{X_i\}$ were smoothed and from the smoothed sequence $\{\tilde{X}_i,\}$ the sequences $\{U_i = \tilde{X}_{i+1}/\tilde{X}_i\}$ and $\{V_i = \log(\tilde{X}_{i+1}/\tilde{X}_i)\}$ were constructed. These were in turn smoothed (to reduce negative correlations between successive elements) yielding $\{\tilde{U}_i\}$ and $\{\tilde{V}_i\}$.

Regression of $\{\tilde{U}_i\}$ on $1/i$ indicated a negative slope while regression of $\{\tilde{V}_i\}$ on $1/i$ gave a slope estimate which when solved for β gave a value of 0.72, quite close to the maximum likelihood estimate. However in neither case was the regression significant, so the evidence in favor of Model B2 is by no means overwhelming. Nonetheless, it suggests that this technique, properly enhanced by smoothing, may be of some value.

We now introduce three quantities which attempt to measure how well the models do at predicting both the IMTBF and the CMTBF. These measures will also allow us to compare the deterministic model with the stochastic ones on a somewhat equal footing. Now stoictly speaking, we should reestimate the parameters of the model after each failure and use these new estimates to form our prediction of the next failure time. This would be rather time consuming, so instead parameters estimated from the entire data set were used to make the predictions. However for the GE1 data the scheme of successive reestimation was carried out and the parameter estimates remained fairly stable, though less so for Model B2.

The statistics described below are closely related to the criteria used in the study described in Schafer, etal [1975]. The first is

$$R1 = \frac{\sum_{i=1}^{n} (X_i-\hat{E}X_i)^2 \;/\; (n-2)}{\sum_{i=1}^{n} (X_i-\overline{X})^2 \;/\; (n-1)} ,$$

where

$$\overline{X} = \sum_{i=1}^{n} X_i/n$$

and

$$\hat{E}\, X_i = \int_0^\infty x d\hat{F}_i(x).$$

The motivation for this is obvious, and we expect R1 to be small (certainly less than one) if the model is of any use. A variant of R1, denoted by R2, is meant to reduce somewhat the effect of the great irregularity in successive interfailure times on R1. We begin at some appropriate time once the process is underway (say after 10 failures) and collect the failure times in successive groups of those after that point. Let

m = number of groups of three

$\tilde{X}_j$ = mean of the observed X_i for the j^{th} group

$\tilde{EX}_j = \hat{E}X_i$ for the middle time in the j^{th} group

$$\bar{\tilde{X}} = \sum_{j=1}^{m} \tilde{X}_j \,/\, m,$$

and define

$$R2 = \frac{\sum_{j=1}^{m} (\tilde{X}_j - E\tilde{X}_j)^2 / (m-2)}{\sum_{j=1}^{m} (\tilde{X}_j - \tilde{X})^2 / (m-1)} .$$

The third criterion, used in Schafer, etal [1975] deals with prediction of CMTBF. Let $\hat{Y}(t_i)$ denote the estimate of $Y(t_i)$, the CMTBF at time t_i. Then

$$R3 = \frac{\sum_{i=1}^{n} \left(Y(t_i) - \hat{Y}(t_i)\right)^2 / (n-2)}{\sum_{i=1}^{n} \left(Y(t_i) - \overline{Y}(t_i)\right)^2 / (n-1)} ,$$

where $\overline{Y}(t_i) = \sum_{i=1}^{n} Y(t_i)/n.$

Table II contains the results of computing these statistics for each of the models and data sets. It seems clear that the stochastic models do better with the IMTBF while the deterministic one does better with the CMTBF, but there is little to choose between the stochastic models. A few general points may be worth mentioning. Since CMTBF is more stable, all the models do well with respect to the R3 criterion, and better with the GE2 than GE1 data. This suggests that looking at only first instances of

failure modes can be one way of obtaining better predictions. Of course there is a practical difficulty that failures can not always be correctly classified or one not even recognized as pattern failures until much later. Secondly, a comparison of the R2 values with R1 values shows that once we have smoothed some of the roughness away, none of the models does better than the sample mean. This is most pronounced with the GE2 data and the inescapable conculsion is that much more insight is required before we can do better. To this end, a simulation was performed which aimed to discover what aspects of the process had the greatest effect on the performance of the models. This is described in the following section.

4. THE SIMULATION

The simulation was designed to produce data that could be reasonably considered to approximate reliability growth data. An observed process of this kind results from the superposition of the failure patterns of different modes (usually due to design faults) together with a process of non pattern failures (wear outs, breakages, etc.) that inevitably occurs during the operating life of any system. Since the GE1 data was used as a guideline, let us review its structure. There are thirteen failure modes which when treated separately as homogeneous poisson processes yield

estimated intensities of between .5 and 2.5 failures per 1000 hrs. (f/th). The non pattern failures have an estimated intensity of 4.4 f/th. The number of observed failures per mode varies between two and seven, with most being either two or three.

Our simulated process will then be constructed as the superposition of a number of independent Poisson processes, each to be truncated after a random number of failures. Even so simple a process defies analytic investigation. It resembles the so-called "branching Poisson process" introduced by Lewis [1964] as a model for computer system failures. His process was stationary but even so, the main results were all asymptotic. For our case we are precisely interested in the transient phase and are not concerned with the structure of the non-pattern failures so asymptotics would by meaningless here. Nonetheless, the work on superposition of independent sparse processes (see the review by Cinlar [1972]) makes it plausible to assume that the superposition will resemble a non-homogeneous Poisson process.

The advantage of this approach is that the generating scheme is based on an analysis of the physical situation and not on a preconceived model. Hence if one of the models does very well, we can have a measure of confidence in its success. Of course with a Monte Carlo we can investigate certain problems which would be difficult if not impossible to do with only real data.

The assumption of independence in the model can not be strictly justified in practice but is probably not a bad approximation to reality. We are primarily interested in the effect of two factors: the number of failure modes and the range of failure intensities. In our simple two-way design each factor has three levels. For the number of failure modes: low (5), medium (10), high (15). The intensity ranges are low, medium, and high, chosen in the following way:

Let m denote the number of failure modes (5,10, or 15). Let $Z_1<Z_2<\ldots<Z_m$ denote the expected values of the order statistics from a uniform (0,1) random sample of size m. The intensities

$\{\lambda_i\}_{i=1}^m$ are chosen so that:

(a) low: $\lambda_i = 2Z_i+.5 \quad (i=1,2,\ldots,m)$

(b) medium $\lambda_i = 2\sqrt{Z_i}+.5 \quad (i=1,2,\ldots,m)$

(c) high $\lambda_i = 2\sqrt[4]{Z_i}+.5 \quad (i=1,2,\ldots,m)$.

(There is clearly overlap among the ranges). We then generate m independent observations $n_1,\ldots,n_m$ from a Poisson distribution with parameter three using the generator described by Ahrens and Dieter [1974]. The n_i are randomly assigned to the λ_i.

Thus we generate m independent Poisson processes, where the i^{th} process has intensity λ_i and is truncated after the appearance of the $n_i{}^{th}$ failure. These are then superposed and to them is added another independent Poisson process with an intensity of 3 f/th, representing the non pattern failures. The result is a simulated reliability growth process which is truncated for fitting and analysis at the first failure after 4500 hours. For each combination of the two factors we have 20 replications, so that there are 3x3x20=180 failure processes generated and analyzed.

It should be emphasized that the twenty replications in a particular cell have the same number of failure modes m, the i^{th} mode having the same intensity λ_i. However the number of failures n_i observed in the i^{th} mode is generated anew for each replication. Of course the waiting times themselves are generated by independent exponential variates. The simulation was programmed in FORTRAN on a PDP 11/40 under the UNIX operating system. A (hopefully!) trustworthy congruential generator was used to produce the uniform deviates employed.

5. ANALYSIS OF RESULTS

Three models were fitted to each replication: D,C, and B2 and the three criteria R1, R2, R3 were calculated for each. (For reasons of economy, Model B1 was jettisoned at this stage.) A discussion and interpretation of the resulting analyses of variance forms the body of this section. Before doing this, it seems useful to mention some of the general impressions gained by use of the graphical techniques introduced in Section 3, though it would be obviously unwieldy to go into too great detail.

First, all the models did better the larger the number of failure modes and the higher the range of intensities - clearly the effect of having more data to work with. The stochastic models had a tendency to underestimate the amount of reliability growth, i.e. overestimating the earlier failure times and underestimating the later ones.

Second, the plots of $\ln Y(t)$ against $\ln t$ often had a U-shaped appearance, initially decreasing but then increasing once the process was well under way. This suggests an interesting modification of Duane's original proposal; namely, estimating the parameters using weighted least squares, with the weights assigned to $\ln Y(t)$ proportional (say) to $\sqrt{N(t)}$. This would certainly give proper due to the increasing number of observations contributing to successive values of $Y(t)$. This proposal was not, however, tried in the present study.

These graphical techniques and even the goodness-of-fit statistics discussed in Section 3 were not sufficiently delicate to discriminate between the models and so we turn to the more quantitative methods. For each criterion (R1,R2,R3) a three factor ANOVA was carried out, employing the criterion as the dependent variable with the factors being models, number of modes, range of intensities.

In addition to the usual ANOVA by least squares fitting, a three-way median polish was carried out (cf. Tukey [1976]). The results were qualitatively the same, though the magnitudes of the fitted effects differed somewhat; there was little indication that a transformation of the data was called for. We therefore content ourselves with presenting only the normal-theory analysis, and Table III contains a compact summary of the results. After considering these, we will discuss some of the more interesting two way tables. The rather large number of degrees of freedom (df) for Residual makes the analyses fairly clearcut.

For R1, the main effects are all highly significant, and the modes-frequency interaction is very significant. Model D does quite a bit worse than C or B2. The differences among the levels within the other two factors are in the expected direction but one not as large in magnitude as for the first. For R2, only models and modes are significant. Note that the residual mean square is

larger than that for R1 by a factor of 300. Again Model D does substantially worse. For R3, only models is significant, but here Model D does substantially better than the others and Model C substantially worse.

As we are led to expect from Section 3, Model D does best with CMTBF, while Models C and B2 do best with IMTBF. It is interesting that as the "smoothness" of the criterion considered increases (R1→R2→R3), the importance of the factors, modes and frequencies, diminishes. Let us investigate this further, by considering some two-way analyses.

Tables IV, V, and VI contain the ANOVAs and fitted effects for R1, R2 and R3 respectively. There we note that Model D is sensitive to the number of modes for R1 and R2 but not R3. Model C is sensitive to the number of modes for all criteria, but to the frequency range only for R1. Model B is sensitive to the number of modes for R1 and R2 and to the frequency range for R1.

However when we consider the data by criterion, we see that Model C tends to be quite a bit more sensitive than the other models to the number of modes, while Models B and D are roughly equivalent in this regard. It is to facilitate this comparison that the tables are arranged as they are. Finally nowhere are the two factor interactions of any importance except possibly with Model C for R1.

6. CONCLUSIONS

There are several conclusions that can be drawn from the study. First the graphical techniques are quite useful in giving a general impression of the fit, but are not delicate enough to distinguish between models. Second, smoothing the data may certainly enhance the stability of the estimation procedures and give some hope of developing ad hoc approaches to model development. The Duane model is quite workable as far as CMTBF is concerned but its performance would probably be improved by use of weighted least squares. It is not recommended when IMTBF is of interest; there the stochastic models are superior. In addition, data of the GE2 type seem to allow regression parameters of the model to be estimated more accurately than data of the GE1 type.

Turning to the simulation, it seems clear that although the various techniques work better with increasing amounts of data, a really successful method would be fairly insensitive to the factors considered such as number of modes and range of intensities. (This is particularly true in the case of R3 which deals with a stable criterion, namely CMTBF.) In fact it is a bit surprising that these factors do not have a greater influence. Of course a larger range of levels ought to be used in any further simulation. Overall Models C and B2 are roughly equivalent in performance, though Model B2 seems less sensitive to changes in the levels of the factors.

We are now in the situation of having the problem somewhat in hand, but by no means mastered. One approach would be to consider stochastic models with different parametizations. The astute reader will have noticed that the models discussed here bear more than a passing resemblance to the linear logistic models of Cox [1970], where it is the log odds ratio that is posited to have a linear regression on one or more independent variables. In the present context it seems easier to work with the discrete $N(t)$ rather than the continuous t itself. (See Brown [1972] for a discussion of hypothesis testing for nonhomogeneous Poisson processes.)

Thus we can envision trying models of the form

$$\log \lambda_i = a+b \log i + c(\log i)^2 \tag{14}$$

trying to get a "best fit". It does not seem likely that this would result in substantial gains. For if one compares the sequence of predicted interfailure times with the actual observed sequence, it is clear that the former, while capturing the drift of the latter, does not resemble it in the least - particularly because of the great variability in the observed times. This situation can not much change by the addition of terms into the regression.

These considerations lead us to suggest that perhaps a doubly stochastic process model might be more suitable (see Grandell [1972].) That is,

$$\log \lambda_i = a+b \log i+\varepsilon_i \tag{15}$$

where $\varepsilon_i \sim N\left(0, \sigma^2\right)$ and independent for each i. Such a model would certainly introduce more variability, but the question arises of whether we understand something better for having named it. The answer in this case is yes, because such a model has ramifications which can be sought in data. If X_i is exponentially distributed with parameter λ_i, then under (14)

$$\mathrm{Var}\ (X_i) = \frac{1}{\lambda_i^2},$$

while under (15)

$$\begin{aligned}\mathrm{Var}\ (X_i) &= E\left[\frac{1}{\lambda_i^2}\right] + \mathrm{Var}\left(\frac{1}{\lambda_i}\right) \\ &\sim \frac{1}{\lambda_i^2} + \mathrm{Var}\left(\frac{1}{\lambda_i}\right).\end{aligned}$$

Thus, in this latter case, there is another component to the variance.

Unfortunately, in any single process we have only one observation on X_i. But it is quite common to test three or four devices simulataneously, thus giving replications which allow us to compute estimated variances. Note that this is only possible when we use N(t) as the independent variable (as in Models B1 and B2) rather than t (as in Model C).

For definiteness, consider making m replicate observations $\left\{X_i(t)\right\}_{j=1}^{m}$ on a particular reliability growth process. We then have n (say) sets of iid observations $\left\{X_{ij} \quad j = 1,2,\ldots,m\right\}_{i=1}^{n}$ where $X_{ij} \sim \mathrm{EXP}\ (\lambda_i)$ $j=1,2,\ldots,m$.

Then compute $\overline{X}_i = \frac{1}{m} \sum_j x_{ij}$ and $s_i^2 = \frac{1}{m-1} \sum_j (X_{ij}-\overline{X}_i)^2$. If we fit a model like (7) to the data (using the full likelihood), then the sequence $\left\{\overline{X}_i\right\}$ should probably conform very closely to $\left\{\frac{1}{\hat{\lambda}_i}\right\}$. But the sequence $\left\{s_i^2\right\}$ would probably not resemble $\left\{\frac{1}{\hat{\lambda}_i^2}\right\}$ very much. If such were the case it would be strong evidence in favor of (15), particularly if the regression of $s_i^2 - \frac{1}{\hat{\lambda}_i^2}$ on i had zero slope. If this regression were not

constant, it would of course indicate modifying (15) to allow the ε_i's to have different variances σ_i^2, and then one might try to study the structure of the σ_i^2. But in this situation there would be also ample incentive to search for another regression which would do as well in fitting the $\bar{X}_i$ while keeping the σ_i^2 constant.

It is certainly premature to proceed further along these lines until supporting evidence for models like (15) can be found. But if they were to prove plausible, they would provide reliability engineers with better variability estimates with which to plan development programs and the like.

Acknowledgments. The authors would like to thank Mr. R.C. Kroeger of the General Electric Company for making available the data analyzed in this paper.

APPENDIX A

For Model B1 we have

$$\lambda(t) = \gamma\beta[N(t)]^{1-1/\beta} .$$

The likelihood function of the observed interfailure times is given by

$$L(x_1,\ldots,x_n) = \prod_{i=1}^{n} \lambda_i \, e^{-\lambda_i x_i}$$

where

$$\lambda_i = \gamma\beta i^{1-1/\beta},$$

taking $N(o) = 1$ for convenience. Differentiating the log likelihood with respect to γ and β and setting the resulting expressions equal to zero yields

$$\hat{\gamma} = \frac{n}{\sum_i \hat{\beta} x_i i^{1-1/\hat{\beta}}} \tag{A1}$$

and

$$\frac{\sum_i i^{1-1/\hat{\beta}} x_i \log i}{\sum_i x_i i^{1-1/\hat{\beta}}} = \frac{\sum_i \log i}{n} . \tag{A2}$$

Equation (A2) can be easily solved numerically for $\hat{\beta}$, the solution being used then in (A1) to obtain $\hat{\gamma}$. Note that various quantities of interest are now easily computed. For example

$E(X_{n+1})$ is estimated by

$$(\hat{\lambda}_{n+1})^{-1} = \left[\hat{\gamma}\hat{\beta}(n+1)^{1-1/\hat{\beta}}\right]^{-1} ,$$

and from this one can calculate the expected number of failures that will occur in a given amount of time. In planning a development program, one may want an estimate of the number of failures that will be needed to bring the reliability to a desired level. If the desired intensity is λ_o and the trial values of the parameters (perhaps current estimates) are γ_o and β_o , then the required number is approximately

$$\left(\frac{\lambda_o}{\gamma_o \beta_o}\right)^{\beta_o/\beta_o - 1}$$

Finally, one would like to have approximate standard errors for the parameter estimates. If we let

$$\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \text{ and } l(\underline{\theta}) = \log L(\underline{\theta}; x_1 \ldots, x_n),$$

then

$$\begin{cases} \dfrac{\partial^2 \, l(\underline{\theta})}{\partial \theta_1^2} = \dfrac{-n}{\theta_1^2} \, , \\[2ex] \dfrac{\partial^2 \, l(\underline{\theta})}{\partial \theta_2^2} \quad \dfrac{-1}{\theta_2^3} \left\{ n\theta_2 + 2 \, \Sigma \log i + \theta_1 \Sigma x_i (\log i)^2 \, i^{1-1/\theta_2} \right\} \, . \\[2ex] \dfrac{\partial^2 \, l(\underline{\theta})}{\partial \theta_1 \partial \theta_2} = -\sum_i x_i \, i^{1-1/\theta_2} \left[1 + \dfrac{1}{\theta_2} \log i \right] \, . \end{cases} \tag{A3}$$

Then as is well known from maximum likelihood theory, (even though the observations are not identically distributed)

$$\left(I_{ij} \; (\underline{\theta}) \right) = \left(E \quad \left[\frac{- \partial^2 \; l(\underline{\theta})}{\partial \theta_i \; \partial \theta j} \right] \right) \, , \tag{A4}$$

which involves replacing x_i in equations (A3) by $\lambda_i^{-1} =$ $\left[\theta_1 \; \theta_2 \; i^{1-1/\theta_2} \right]^{-1}$. . The resulting expression depends on the data only through $\hat{\theta}_1$, $\hat{\theta}_2$ and n, and is particularly simple to compute. The asymptotic variance-covariance matrix of $\hat{\theta}$ is the

inverse matrix to (A4), denoted by $\left(I^{ij}(\underline{\theta})\right)$ and which has for a consistent estimate, $\left(I^{ij}(\hat{\underline{\theta}})\right)$.

For Model B2 we have

$$\lambda_i = c \exp\left\{b\left(1-\frac{i-1}{n}\right)^2\right\}.$$

Differentiating the log likelihood with respect to b and c and setting the resulting expressions equal to zero yields

$$\hat{c} = n/\sum_i x_i \exp\left\{\hat{b}\left(1-\frac{i-1}{n}\right)^2\right. \tag{A5}$$

and

$$\left[\sum_i x_i\left\{\exp\ \hat{b}\left(1-\frac{i-1}{n}\right)^2\right\}\right]\left[\sum_i\left(1-\frac{i-1}{n}\right)^2\right]\Big/ n \tag{A6}$$

$$= \sum_i x_i \exp\left\{\hat{b}\left(1-\frac{i-1}{n}\right)^2\right\}\left(1-\frac{i-1}{n}\right)^2.$$

Equation (A6) can be easily solved numerically for $\hat{b}$, the solution being used in (A5) to obtain $\hat{c}$.

Now let $\underline{\theta} = \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \begin{pmatrix}c\\ b\end{pmatrix}$ and

$l(\underline{\theta}) = \log L(\underline{\theta};\ x_1,\ldots,x_n)$. Then

$$
\left\{
\begin{aligned}
\frac{\partial^2 l(\underline{\theta})}{\partial \theta_1^2} &= -\frac{n}{\theta_1^2}, \\
\frac{\partial^2 l(\underline{\theta})}{\partial \theta_2^2} &= -\theta_1 \sum_i x_i \exp\left\{\theta_2 \left(1-\frac{i-1}{n}\right)^2\right\}\left(1-\frac{i-1}{n}\right)^4, \\
\frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_2} &= -\sum_i x_i \exp\left\{\theta_2 \left(1-\frac{i-1}{n}\right)^2\right\}\left(1-\frac{i-1}{n}\right)^2.
\end{aligned}
\right. \tag{A7}
$$

The argument after equations (A3) applies here as well, and a consistent estimate of the variance-covariance matrix of $\underline{\theta}$ is obtained. In this case the variance-covariance matrix depends only on $\hat{\theta}_1$, and n.

APPENDIX B

A. GE1 Data. Interfailure Times (to be read across rows)

25	15	210	25	20	5	15	30	5	25
75	10	110	10	80	120	60	110	10	60
30	25	175	175	25	200	175	25	10	65
25	250	5	95	50	25	45	5	15	60
70	10	170	20	80	30	195	125	100	150
60	190								

B GE2 Data. Interfailure Times (to be read across rows)

25(2)	15(NP)	210(NP)	25(NP)	20(3)
5(2)	15(NP)	30(7)	5(NP)	25(2)
85(NP)	110(NP)	10(NP)	260(2)	180(2)
55(NP)	175(2)	175(NP)	500(1)	25(3)
255(NP)	215(3)	5(NP)	15(5)	130(NP)
200(NP)	740(NP)			

NOTE: Quantities in brackets give the number of failures that occurred in that mode whose first occurrance is recorded in the corresponding entry. "NP" stands for non-pattern failure.

TABLE 1: PARAMETER ESTIMATES

Model \ Data	GE1	GE2
D	$\hat{b}$ = .195 $\hat{a}$ = 2.65	$\hat{b}$ = .347 $\hat{a}$ = 1.89
C	$\hat{\beta}$ = .837 $\hat{\gamma}$ = .0537	$\hat{\beta}$ = .620 $\hat{\gamma}$ = .171
B1	$\hat{\beta}$ = .799 (.1002) $\hat{\gamma}$ = .0385 (.0236)	$\hat{\beta}$ = .592 (.00817) $\hat{\gamma}$ = .0799 (.00577)
B2	$\hat{b}$ = .635 (.457) $\hat{c}$ = .0121 (.00253)	$\hat{b}$ = 1.49 (.0197) $\hat{c}$ = .0004 (.000117)

NOTE: Quantities in brackets are estimated standard errors of estimates.

TABLE 2: GOODNESS-OF-FIT STATISTICS

ON MODEL PREDICTIONS

	R1		R2		R3	
MODEL	GE1	GE2	GE1	GE2	GE1	GE2
D	.986	.921	1.18	1.96	.454	.179
C	.983	.870	1.19	1.55	.480	.228
B1	.963	.833	1.17	1.41	.621	.472
B2	.969	.894	1.22	1.94	.571	.322

TABLE 3: 3-FACTOR ANOVA'S AND ESTIMATED MAIN EFFECTS (FROM SIMULATION)

Source of Variation		df	R1 F-Stat	R1 Est'd Effect	R2 F-Stat	R2 Est'd Effect	R3 F-Stat	R3 Est'd Effect
Grand Mean		1	-	.92	-	1.37	-	.52
Models	D	2	76.8***	.07	5.74**	.35	13.1***	-.13
	B2			-.04		-.09		.01
	C			-.03		-.26		.12
Modes	5	2	18.8***	.04	7.85***	.43	2.37	.06
	10			-.01		-.20		-.05
	15			-.02		-.23		-.01
Freq	Low	2	10.1***	.02	.05	-.01	1.36	.04
	Med			.0		-.03		-.04
	High			-.02		.03		.0
Models x Modes		4	.77		.96		1.62	
Models x Freq		4	1.27		1.03		1.40	
Modes x Freq		4	3.18*		.04		1.44	
Models x Modes x Freq		8	.16		.76		.81	
			MS		MS		MS	
Residual		513	(.010)		(3.189)		(.204)	
Total		540						

NOTES: (1) Significance levels of F-statistics: *→p<.05; **→p<.01, ***→p<.001

(2) Bracketed quantities in "Residual" row are mean squares. Thus all mean squares and sums of squares can be computed from information in Table

(3) e.g. Estimated main effect of B2 with R1 as dependent variable is .92+(-.04)=.88

TABLE 4: 2-FACTOR ANOVA'S WITH R1 AS DEPENDENT VARIABLE; ALSO ESTIMATED MAIN EFFECTS

Source of Variation		df	Model D F-Stat	Model D Est'd Effect	Model C F-Stat	Model C Est'd Effect	Model B2 F-Stat	Model B2 Est'd Effect
Grand Mean		1	-	1.0	-	.89	-	.88
Modes	5			.04		.04		.03
	10	2	5.36**	-.02	10.72***	-.01	6.48**	-.02
	15			-.02		-.03		-.01
Freq	Low			.01		.04		.03
	Med	2	.36	-.01	10.45***	-.01	7.17***	.0
	High			-.01		-.03		-.03
Modes x Freq		4	.83		2.19		1.03	
			MS		MS		MS	
Residual		171	(.016)		(.007)		(.007)	
Total		180						

See Notes to Table 3

TABLE 5: 2-FACTOR ANOVA'S WITH R2 AS DEPENDENT VARIABLE; ALSO ESTIMATED MAIN EFFECTS

Source of Variation		df	MODEL D F-Stat	MODEL D Est'd Effect	MODEL C F-Stat	MODEL C Est'd Effect	MODEL B2 F-Stat	MODEL B2 Est'd Effect
Grand Mean		1	-	1.72	-	1.10	-	1.28
Modes	5			.70		.19		.40
	10	2	3.17*	-.37	7.04**	-.09	3.14	-.15
	15			-.33		-.10		-.25
Freq	Low			-.27		-.03		.27
	Med	2	.46	.10	.67	-.03	1.44	-.15
	High			.17		.06		-.12
Modes x Freq		4	.18		.85		1.48	
			MS		MS		MS	
Residual		171	(7.02)		(.234)		(2.31)	
Total		180						

TABLE 6: 2 FACTOR ANOVA'S WITH R3 AS DEPENDENT VARIABLE

Source	df	D	C	B2
Grand Mean	1	-	-	-
Modes	2	.24	3.45*	.05
Freq	2	1.66	1.96	.13
Modes x Freq	4	1.02	.94	1.16
		MS	MS	MS
Residual	171	(.120)	(.321)	(.173)